

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:36 ; Search time 30.11 Seconds
(without alignments)
464.806 Million cell updates/sec

Title: US-09-863-063-2
Perfect score: 682
Sequence: 1 AQSVPFGDIQTGPKIVFN.....EWFQDGMVRKKNLPPIEYNP 126

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	16.7	242	21 AAY94850	Human protein clon
2	113	16.6	236	20 AAW86813	Human vesicular bl
3	113	16.6	300	20 AAY76635	Human ovarian tumo
4	105.5	15.5	341	22 AA012925	Human polypeptide
5	101	14.8	626	22 ABG01123	Novel human diagno
6	99	14.5	308	22 AAG03012	Novel human secret
7	98	14.4	76	22 AAU20373	Human secreted pro
8	95	13.9	223	22 AAU01095	Gene 27 Human secr
9	95	13.9	243	20 AAY41727	Human PRO983 prote
10	95	13.9	243	21 AAB44283	Human PRO983 (UNQ4
11	95	13.9	243	22 AAU29079	Human PRO polypept

12	95	13.9	243	22	AAB95358	Human protein sequ
13	85	12.5	16	18	AAW31324	Dictyocaulus vivip
14	83	12.2	204	21	AAG48788	Arabidopsis thalia
15	83	12.2	266	21	AAG48787	Arabidopsis thalia
16	81	11.9	293	21	AAG07466	Arabidopsis thalia
17	81	11.9	293	21	AAG48106	Arabidopsis thalia
18	81	11.9	295	21	AAG07465	Arabidopsis thalia
19	81	11.9	295	21	AAG48105	Arabidopsis thalia
20	79	11.6	220	21	AAG08266	Arabidopsis thalia
21	79	11.6	220	21	AAG14886	Arabidopsis thalia
22	79	11.6	220	21	AAG48818	Arabidopsis thalia
23	79	11.6	220	21	AAG48858	Arabidopsis thalia
24	78.5	11.5	269	22	ABB61176	Drosophila melanog
25	76	11.1	20	18	AAW31322	Dictyocaulus vivip
26	75.5	11.1	201	22	ABB63589	Drosophila melanog
27	73	10.7	112	22	AAU30211	Novel human secret
28	72	10.6	166	21	AAG07157	Arabidopsis thalia
29	72	10.6	166	21	AAG48809	Arabidopsis thalia
30	72	10.6	202	21	AAG36223	Arabidopsis thalia
31	72	10.6	250	21	AAG36222	Arabidopsis thalia
32	72	10.6	287	21	AAG36221	Arabidopsis thalia
33	72	10.6	1069	22	ABB63364	Drosophila melanog
34	71.5	10.5	363	22	ABB63365	Drosophila melanog
35	70.5	10.3	471	22	AAU33719	Pseudomonas aerugi
36	69.5	10.2	4618	22	AAW39043	Human polypeptide
37	68	10.0	294	22	AAG76672	Corynebacterium gl
38	68	10.0	529	22	AAG91430	C glutamicum prote
39	68	10.0	787	22	AAB96500	Putative P. abyss
40	67.5	9.9	826	22	ABG21280	Novel human diagno
41	67	9.8	414	22	ABB67911	Drosophila melanog
42	67	9.8	414	22	AAU38970	Drosophila C-prote
43	66.5	9.8	144	22	AAW80439	Gene #20 associate
44	66.5	9.8	163	22	AAB80438	Gene #20 associate
45	66	9.7	248	21	AAG09870	Arabidopsis thalia

ALIGNMENTS

RESULT	1
AA194850	
ID	AA194850 standard; Protein; 242 AA.
XX	AC
XX	AA194850;
XX	AC
XX	12-JUN-2000 (first entry)
XX	Human protein clone HP10195.
DE	Human protein; hydrophobic domain; nutritional source; haematopoiesis;
XX	cytokine production; cell proliferation; cell differentiation;
KW	immune deficiency; infectious disease; autoimmune disorder; asthma;
KW	multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KW	allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
KW	nervous system disorder; Alzheimer's disease; Parkinson's disease;
KW	Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
KW	systemic cytokine damage; tissue differentiation; contraceptive; stroke;
KW	coagulation disorder; myocardial infarction; inflammatory condition;
KW	septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
KW	nephritis; therapy.
XX	Homo sapiens.
XX	OS
XX	WO200005367-A2.
PN	XX
XX	03-FEB-2000.
PD	XX
XX	22-JUL-1999; 99WO-JP03929.
PF	XX
XX	24-JUL-1998; 98JP-0208820.
PR	07-AUG-1998; 98JP-0224105.
PR	25-AUG-1998; 98JP-0238116.
PR	09-SEP-1998; 98JP-0254736.

PD 21-OCT-1999.
 XX
 PF 09-APR-1998; 98DE-1017557.
 XX
 PR 09-APR-1998; 98DE-1017557.
 XX
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;
 XX
 DR WPI; 1999-591920/51.
 DR N-PSDB; AAZ77504.
 XX
 PT New nucleic acid sequences expressed in ovarian, and some other, cancer
 PT tissues, and derived polypeptides, for treatment of ovarian cancer and
 PT identification of therapeutic agents -
 XX
 XX Claim 25; Page 298; 310pp; German.
 XX
 CC This invention describes novel nucleic acid (cDNA) sequences (A) which
 CC have anticancer activity and are highly expressed in ovarian tumor
 CC tissue (and some also in testis and breast cancer tissue). The products
 CC of the invention can be used for gene therapy. (A) are used (i) for
 CC recombinant expression of polypeptides (B) and (ii) to isolate complete
 CC genes. (B) are used (i) to identify agents suitable for treatment of
 CC ovarian cancer; (ii) directly for treating this form of cancer
 CC (including expression from gene therapy vectors) and (iii) for generation
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed
 CC sequence tags) from a particular tissue type before comparison of
 CC expression patterns. This allows a significantly longer fragment of the
 CC gene to be revealed, so should reduce the number of failures associated
 CC with the fact that ESTs from different libraries may represent different
 CC parts of the same unknown gene, distorting the estimated frequency of
 CC occurrence in a particular tissue. AAY76505-Y76638 represent protein
 CC fragments encoded by the human ovarian tumor cDNA library derived EST
 CC fragments represented in AAZ77450-277572.
 XX
 XX Sequence 300 AA;
 SQ
 Query Match 16.6%; Score 113; DB 20; Length 300;
 Best Local Similarity 28.2%; Pred. No. 2.3e-05;
 Matches 24; Conservative 19; Mismatches 42; Indels 0; Gaps 0;
 QY 13 PGKIVFNAPYDDKHTYHKVINSSARRIGYGIKTNNMKRLGVDPGCVLDPKKAVLLAV 72
 DB 70 pptdlkfkgpfdvvttnklrpsdrkvcfkvtktaprycvcvprnsgiidpgstvtvs 129
 QY 73 SCDAFAFGQEDNNDRIIVENTNP 97
 DB 130 mlqffdydpnekskhkfmvgtifap 154
 RESULT 4
 ID AAO12925
 AC AAO12925 standard; Protein; 341 AA.
 AC AAO12925;
 XX
 XX 06-NOV-2001 (first entry)
 DT
 DE Human polypeptide SEQ ID NO 26817.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 XX WO200164835-A2.
 PN
 XX 07-SEP-2001.
 PR

XX 26-FEB-2001; 2001WO-US04927.
 PF
 XX 28-FEB-2000; 2000US-0515126.
 PR
 XX 18-MAY-2000; 2000US-0577409.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX WPI; 2001-514838/56.
 DR
 DR N-PSDB; AAI92856.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX
 XX Claim 20; SEQ ID NO 26817; 1399pp + Sequence Listing; English.
 PS
 XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 341 AA;
 SQ
 Query Match 15.5%; Score 105.5; DB 22; Length 341;
 Best Local Similarity 27.8%; Pred. No. 0.00024;
 Matches 27; Conservative 19; Mismatches 40; Indels 11; Gaps 2;
 QY 5 PFEDIQOTPGTKIVFNAPYDDKHTYHKVINSSARRIGYGIKTNNMKRLGVDPGCVLD 64
 DB 52 pptdlkfkgpfdvvttnklrpsdrkvcfkvtktaprycvcvprnsgiidpgstvtvs 103
 QY 65 KEAVLLAVSCDAFAFGQEDNNDRIIVENTNP 98
 DB 104 gstvsvmlqpfaydpnekskhkfmvgtifap 140
 RESULT 5
 ID ABG01123
 AC ABG01123 standard; Protein; 626 AA.
 AC ABG01123;
 XX
 XX 13-FEB-2002 (first entry)
 DT
 DE Novel human diagnostic protein #1114.
 XX
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR
 XX 23-AUG-2000; 2000US-0649167.
 PR

PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX (GETH) GENENTECH INC.
 XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI; 2001-602746/68.
 DR N-PSDB; AAS45980.
 XX
 PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 PS Claim 11; Fig 112; 774pp; English.
 XX
 CC Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
 CC The PRO polypeptides and their associated nucleic acids can be used to
 CC detect the presence of a tumour in a mammal by comparing the level of
 CC expression of a PRO polypeptide in a test sample of cells from the animal
 CC and a control sample of normal cells, whereby a higher level of
 CC expression in the test sample indicates the presence of a tumour in the
 CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
 CC and rabbits but are preferably human. The polypeptides can be used to
 CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
 CC when contacted with it. A specific polypeptide can be used to stimulate
 CC the proliferation or differentiation of chondrocyte cells. The PRO
 CC proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 XX
 SQ Sequence 243 AA;

 Query Match 13.9%; Score 95; DB 22; Length 243;
 Best Local Similarity 23.4%; Pred. No. 0.003;
 Matches 22; Conservative 25; Mismatches 43; Indels 4; Gaps 1;

 QY 9 IQTQPGTKIVFNAPYDDKHVYHIKVINSSARRIGYGIKTTNMKRLGVDPGVLDPKEAV 68
 Db 8 lslepqhelkfrgftdvvtlnklgnptdrnvcfkvtaprrycvrpnsgildagasi 67

 QY 69 LLAVSCDAFAFGQEDTNNDRITVEW---TNTPD 98
 Db 68 nvsvmlqpfdydpnekskhkfmvqsmfapttdtsd 101

 RESULT 12
 AAB95358
 ID AAB95358 standard; Protein; 243 AA.
 XX
 AC AAB95358;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17658.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX

PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HELI-) HELIX RES INST.
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 17658; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 243 AA;

 Query Match 13.9%; Score 95; DB 22; Length 243;
 Best Local Similarity 23.4%; Pred. No. 0.003;
 Matches 22; Conservative 25; Mismatches 43; Indels 4; Gaps 1;

 QY 9 IQTQPGTKIVFNAPYDDKHVYHIKVINSSARRIGYGIKTTNMKRLGVDPGVLDPKEAV 68
 Db 8 lslepqhelkfrgftdvvtlnklgnptdrnvcfkvtaprrycvrpnsgildagasi 67

 QY 69 LLAVSCDAFAFGQEDTNNDRITVEW---TNTPD 98
 Db 68 nvsvmlqpfdydpnekskhkfmvqsmfapttdtsd 101

 RESULT 13
 AAW31324
 ID AAW31324 standard; peptide; 16 AA.
 XX
 AC AAW31324;
 XX
 DT 13-MAR-1998 (first entry)
 XX
 DE Dictyocaulus viviparus DV18 peptide fragment 4.
 XX
 KW Immunogenic protein; DV18; lungworm; vaccine; immunity;
 KW dictyocauliasis; cattle; immunoassay.
 XX
 PF Dictyocaulus viviparus.
 OS


```

XX EP785253-A1.
PN
XX
XX 23-JUL-1997.
PD
XX
XX 27-DEC-1996; 96EP-0120947.
PF
XX
XX 19-JAN-1996; 96DE-4001754.
PR
XX
XX (FARH ) HOECHST AG.
PA
XX
XX Hofmann J, Schmid K;
PI
XX
XX WPI; 1997-365928/34.
DR
XX
XX Immunogenic Dictyocaulus viviparus lungworm protein - for use in
PT vaccines and immunoassays
XX
XX Example 10; Page 6; 17pp; German.
PS
XX
XX This sequence represents a peptide fragment of the immunogenic protein
CC DV18 isolated from adult lungworms. The N-terminal of this amino acid
CC can be joined to the amino acids Asp, Asp, Ser, Gly or Arg. DV18 can be
CC used in vaccines for immunising cattle against dictyocauliasis and in an
CC ELISA immunoassay for determining DV18-specific antibodies in the blood
CC of cattle.
XX
XX Sequence 16 AA;
SQ

Query Match 12.5%; Score 85; DS 18; Length 16;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 FRWFQGDGMVRRK 118
Db 2 frwfgdgmvrk 16

RESULT 14
AAG48788
ID AAG48788 standard; Protein; 204 AA.
XX
XX AAG48788;
AC
XX
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 61647.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR
XX
XX 05-MAR-1999; 99US-0123180.
PR
XX
XX 09-MAR-1999; 99US-0123548.
PR
XX
XX 23-MAR-1999; 99US-0125788.
PR
XX
XX 25-MAR-1999; 99US-0126264.
PR
XX
XX 29-MAR-1999; 99US-0126785.
PR
XX
XX 01-APR-1999; 99US-0127462.
PR
XX
XX 06-APR-1999; 99US-0128234.
PR
XX
XX 08-APR-1999; 99US-0128714.
PR
XX
XX 16-APR-1999; 99US-0129845.
PR
XX
XX 19-APR-1999; 99US-0130077.
PR
XX
XX 21-APR-1999; 99US-0130449.
PR
XX
XX 23-APR-1999; 99US-0130510.
PR
XX
XX 28-APR-1999; 99US-0130891.
PR
XX
XX 30-APR-1999; 99US-0131449.
PR
XX
XX 04-MAY-1999; 99US-0132407.
PR
XX
XX 05-MAY-1999; 99US-0132484.
PR
XX
XX 06-MAY-1999; 99US-0132486.
PR
XX
XX 07-MAY-1999; 99US-0132487.
PR
XX
XX 11-MAY-1999; 99US-0132863.
PR
XX
XX 14-MAY-1999; 99US-0134256.
PR
XX
XX 14-MAY-1999; 99US-0134218.
PR
XX
XX 14-MAY-1999; 99US-0134219.
PR
XX
XX 14-MAY-1999; 99US-0134221.
PR
XX
XX 18-MAY-1999; 99US-0134370.
PR
XX
XX 19-MAY-1999; 99US-0134768.
PR
XX
XX 20-MAY-1999; 99US-0134941.
PR
XX
XX 21-MAY-1999; 99US-0135124.
PR
XX
XX 24-MAY-1999; 99US-0135353.
PR
XX
XX 25-MAY-1999; 99US-0135629.
PR
XX
XX 27-MAY-1999; 99US-0136021.
PR
XX
XX 28-MAY-1999; 99US-0136392.
PR
XX
XX 01-JUN-1999; 99US-0136782.
PR
XX
XX 03-JUN-1999; 99US-0137222.
PR
XX
XX 04-JUN-1999; 99US-0137528.
PR
XX
XX 07-JUN-1999; 99US-0137502.
PR
XX
XX 08-JUN-1999; 99US-0137724.
PR
XX
XX 10-JUN-1999; 99US-0138094.
PR
XX
XX 10-JUN-1999; 99US-0138540.
PR
XX
XX 14-JUN-1999; 99US-0138847.
PR
XX
XX 16-JUN-1999; 99US-0139119.
PR
XX
XX 16-JUN-1999; 99US-0139452.
PR
XX
XX 16-JUN-1999; 99US-0139453.
PR
XX
XX 17-JUN-1999; 99US-0139492.
PR
XX
XX 18-JUN-1999; 99US-0139454.
PR
XX
XX 18-JUN-1999; 99US-0139455.
PR
XX
XX 18-JUN-1999; 99US-0139456.
PR
XX
XX 18-JUN-1999; 99US-0139457.
PR
XX
XX 18-JUN-1999; 99US-0139458.
PR
XX
XX 18-JUN-1999; 99US-0139459.
PR
XX
XX 18-JUN-1999; 99US-0139460.
PR
XX
XX 18-JUN-1999; 99US-0139461.
PR
XX
XX 18-JUN-1999; 99US-0139462.
PR
XX
XX 18-JUN-1999; 99US-0139463.
PR
XX
XX 18-JUN-1999; 99US-0139750.
PR
XX
XX 21-JUN-1999; 99US-0139763.
PR
XX
XX 22-JUN-1999; 99US-0139817.
PR
XX
XX 23-JUN-1999; 99US-0139899.
PR
XX
XX 23-JUN-1999; 99US-0140353.
PR
XX
XX 23-JUN-1999; 99US-0140354.
PR
XX
XX 24-JUN-1999; 99US-0140695.
PR
XX
XX 28-JUN-1999; 99US-0140823.
PR
XX
XX 29-JUN-1999; 99US-0140991.
PR
XX
XX 30-JUN-1999; 99US-0141287.
PR
XX
XX 01-JUL-1999; 99US-0141842.
PR
XX
XX 01-JUL-1999; 99US-0142154.
PR
XX
XX 02-JUL-1999; 99US-0142055.
PR
XX
XX 08-JUL-1999; 99US-0142390.
PR
XX
XX 08-JUL-1999; 99US-0142803.
PR
XX
XX 09-JUL-1999; 99US-0142920.
PR
XX
XX 12-JUL-1999; 99US-0142977.
PR
XX
XX 13-JUL-1999; 99US-0143542.
PR
XX
XX 14-JUL-1999; 99US-0143624.
PR
XX
XX 15-JUL-1999; 99US-0144005.
PR
XX
XX 16-JUL-1999; 99US-0144085.
PR
XX
XX 16-JUL-1999; 99US-0144086.
PR
XX
XX 19-JUL-1999; 99US-0144325.
PR
XX
XX 19-JUL-1999; 99US-0144331.
PR
XX
XX 19-JUL-1999; 99US-0144332.
PR
XX
XX 19-JUL-1999; 99US-0144333.
PR
XX
XX 19-JUL-1999; 99US-0144334.
PR
XX
XX 19-JUL-1999; 99US-0144335.
PR
XX
XX 20-JUL-1999; 99US-0144352.
PR

```

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0132866.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 29-JUN-1999; 99US-0141287.
PR 30-JUN-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

Search completed: September 20, 2002, 09:23:16
Job time: 40 sec.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:42 ; Search time 25.65 Seconds
(without alignments)
119.985 Million cell updates/sec

Title: US-09-863-063-2
Perfect score: 682
Sequence: 1 AQSVPGGDIQTQPKTKVFN.....ENFQGDGMVRRKLNLPTEYNP 126

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2.6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2.6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2.6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2.6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2.6/ptodata/2/1aa/PTCUS.COMB.pep.*
6: /cgn2.6/ptodata/2/1aa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	593	87.0	127	4	US-08-786-455B-14 Sequence 14, Appl
2	134	19.6	260	3	US-08-857-213-1 Sequence 3, Appl
3	113	16.6	236	3	US-08-857-213-1 Sequence 1, Appl
4	76	11.1	20	4	US-08-786-455B-2 Sequence 2, Appl
5	66	9.7	1003	1	US-08-571-758-4 Sequence 4, Appl
6	66	9.7	1003	1	US-08-909-984A-4 Sequence 4, Appl
7	66	9.7	1003	1	US-08-909-983-4 Sequence 4, Appl
8	65.5	9.6	890	2	US-08-483-101-14 Sequence 14, Appl
9	64.5	9.5	14	4	US-08-786-455B-4 Sequence 4, Appl
10	63.5	9.3	305	2	US-08-853-659A-47 Sequence 47, Appl
11	62.5	9.2	139	2	US-08-553-501A-29 Sequence 29, Appl
12	62.5	9.2	139	3	US-09-205-231-29 Sequence 29, Appl
13	62	9.1	1076	2	US-08-867-941-19 Sequence 19, Appl
14	62	9.1	1076	4	US-09-074-658-19 Sequence 19, Appl
15	61.5	9.0	405	5	PCT-US93-11404-2 Sequence 2, Appl
16	61	8.9	995	2	US-08-673-789-5 Sequence 5, Appl
17	61	8.9	1013	2	US-08-866-650-3 Sequence 3, Appl
18	61	8.9	1013	2	US-09-021-287-3 Sequence 3, Appl
19	61	8.9	1013	4	US-09-240-473-3 Sequence 3, Appl
20	61	8.9	1864	2	US-08-804-227C-3 Sequence 3, Appl
21	60.5	8.9	120	2	US-08-553-501A-88 Sequence 88, Appl
22	60.5	8.9	120	3	US-09-205-231-88 Sequence 88, Appl
23	60.5	8.9	312	4	US-09-475-316A-48 Sequence 48, Appl
24	60.5	8.9	312	4	US-09-475-316A-50 Sequence 50, Appl
25	60.5	8.9	312	4	US-09-475-316A-52 Sequence 52, Appl
26	60.5	8.9	312	4	US-09-475-316A-54 Sequence 54, Appl
27	60.5	8.9	312	4	US-09-475-316A-56 Sequence 56, Appl

28	60.5	8.9	312	4	US-09-475-316A-58 Sequence 58, Appl
29	60.5	8.9	979	3	US-08-870-529-2 Sequence 2, Appl
30	60	8.8	113	2	US-08-256-568B-83 Sequence 83, Appl
31	60	8.8	113	4	US-09-038-369B-83 Sequence 83, Appl
32	59.5	8.7	246	1	US-08-197-834-7 Sequence 7, Appl
33	59.5	8.7	735	2	US-08-313-185-48 Sequence 48, Appl
34	59.5	8.7	735	2	US-08-459-499-9 Sequence 9, Appl
35	59.5	8.7	735	2	US-08-459-499-12 Sequence 12, Appl
36	59.5	8.7	735	3	US-09-082-614A-48 Sequence 48, Appl
37	59.5	8.7	1288	3	US-08-762-428A-6 Sequence 6, Appl
38	58	8.5	897	1	US-08-095-737-4 Sequence 4, Appl
39	58	8.5	897	1	US-08-480-145-4 Sequence 4, Appl
40	58	8.5	897	2	US-08-477-389-4 Sequence 4, Appl
41	57.5	8.4	172	2	US-08-853-659A-46 Sequence 46, Appl
42	57.5	8.4	249	4	US-09-345-468-9 Sequence 9, Appl
43	57.5	8.4	319	4	US-09-345-468-5 Sequence 5, Appl
44	57.5	8.4	339	4	US-09-345-468-3 Sequence 3, Appl
45	57.5	8.4	654	1	US-08-392-828C-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-786-455B-14
; Sequence 14, Application US/08786455B
; Patent No. 6193971
; GENERAL INFORMATION:
; APPLICANT: HOFMANN, Joachim
; TITLE OF INVENTION: DICTYOCALUS VIVIPARUS ANTIGEN FOR
; TITLE OF INVENTION: DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,455B
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 01 754.8
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/327
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 127 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-786-455B-14

Query Match 87.0%; Score 593; DB 4; Length 127;
Best Local Similarity 86.2%; Pred. No. 1 le-67;
Matches 106; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 3 SVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCGVL 62
Db 3 SVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCGVL 62
QY 63 DPKEAVLLVSCDAFAFGQEDTNNDRITVWNTPDGAARQFRFREWFGDGMVRRKKNLPI 122
Db 63 DPKEAVLLVSCDAFAFGQEDTNNDRITVWNTPDGAARQFRFREWFGDGMVRRKKNLPI 122
QY 123 EYN 125
Db 123 EYN 125

RESULT 2
US-08-857-213-3
; Sequence 3, Application US/08857213
; Patent No. 6054290
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN VESICLE BINDING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/857,213
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0297 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 260 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1000369
US-08-857-213-3

Query Match 19.6%; Score 134; DB 3; Length 260;
Best Local Similarity 29.9%; Pred. No. 6e-09; Mismatches 50; Indels 0; Gaps 0;
Matches 29; Conservative 18;
QY 12 QPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCGVLDPKEAVLLA 71
Db 11 EPAGELRFKPGFTDVTADIKLSNPDRICEFKVKTAPKRYCVRPNSGILEPKTSIAVA 70
QY 72 VSCDAFAFGQEDTNNDRITVWNTPDGAARQFRFREW 108
Db 71 VM-LQPFNYDPNEKNKHKFMVQSMYAPDHHVVSQELLW 107

RESULT 3
US-08-857-213-1
; Sequence 1, Application US/08857213
; Patent No. 6054290
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN VESICLE BINDING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/857,213
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0297 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: FIBRNGT01
; CLONE: 148415
US-08-857-213-1

Query Match 16.6%; Score 113; DB 3; Length 236;
Best Local Similarity 28.8%; Pred. No. 2.4e-06;
Matches 23; Conservative 19; Mismatches 38; Indels 0; Gaps 0;
QY 13 PGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCGVLDPKEAVLLAV 72
Db 12 PPTDLAFKPGFTDVTADIKLSNPDRICEFKVKTAPKRYCVRPNSGILEPKTSIAVA 71
QY 73 SCDAFAFGQEDTNNDRITVE 92
Db 72 MLQPFYDNEKNKHKFMVQ 91

RESULT 4
US-08-786-455B-2
; Sequence 2, Application US/08786455B
; Patent No. 6193971
; GENERAL INFORMATION:
; APPLICANT: HOFMANN, Joachim
; APPLICANT: SCHMID, Karlheinz
; TITLE OF INVENTION: DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION

```

; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,455B
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 01 754.8
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/327
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-786-455B-2

```

```

Query Match 11.1%; Score 76; DB 4; Length 20;
Best Local Similarity 78.9%; Pred. No. 0.0036;
Matches 15; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 53 LGVDPPCGVLDPKREAVLLA 71
Db 2 LGVDPPASGVLDPKREATLMA 20

```

```

RESULT 5
; Sequence 4, Application US/08571758
; Patent No. 5700675
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerry M.
; APPLICANT: Therrien, Marc
; APPLICANT: Chang, Henry C.
; APPLICANT: Karim, Felix D.
; APPLICANT: Wassarman, David A.
; TITLE OF INVENTION: A No. 5700675el Protein Kinase Required for Ras
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/571,758

```

```

; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B96-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1003 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; US-08-571-758-4

```

```

Query Match 9.7%; Score 66; DB 1; Length 1003;
Best Local Similarity 21.0%; Pred. NO. 18;
Matches 26; Conservative 19; Mismatches 45; Indels 34; Gaps 4;

```

```

QY 3 SYPPGGDIQTQPGTKIVENAPYDDKHTYHIKVINSSA-----RRIGYGIKTTNM 50
Db 602 SLTGSQVSTHSATSQSVNSVSGSSSATYTSLLVNSGFFPRKLSNAGVDRKVPFTSEYTD 661
QY 51 KRLGVDPGCVLDPKREAVLLAVSCDAFAFQGEDTNDRIITVETWTPDGAARQFR----- 105
Db 662 HKSN-----DSDKTVSL-----GSASTDSRTPVRLDSTEDGDSGQWRQNSIS 705
QY 106 -REM 108
Db 706 LKEW 709

```

```

RESULT 6
; Sequence 4, Application US/08909984A
; Patent No. 5747275
; GENERAL INFORMATION:
; APPLICANT: Rubin, Gerry M.
; APPLICANT: Therrien, Marc
; APPLICANT: Chang, Henry C.
; APPLICANT: Karim, Felix D.
; APPLICANT: Wassarman, David A.
; TITLE OF INVENTION: A No. 5747275el Protein Kinase Required for Ras
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 268 BUSH STREET, SUITE 3200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909,984A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: B96-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 343-4341
; TELEFAX: (415) 343-4342
; INFORMATION FOR SEQ ID NO: 4:

```

```
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1003 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: peptide
;
US-08-909-984A-4

Query Match          9.7%; Score 66; DB 1; Length 1003;
Best Local Similarity 21.0%; Pred. No. 18;
Matches 26; Conservative 19; Mismatches 45; Indels 34; Gaps 4;

QY 3 SVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSA-----RRIGYGIKTTNM 50
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 602 SLTGSQVSTHSATSQVNSVSGSSSATYTSLLVNSGFFPRKLSNAGVDKRVFFSEYTD 661
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 51 KRLGVDPGCVLDPKRAVLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFR----- 105
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 662 HKSN-----DSDKTVSLS-----GSASTDSDRTPVRLDSTEDGDSGQWRQNSIS 705
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 106 -REW 108
   :||
Db 706 LKEW 709

RESULT 7
US-08-909-983-4
; Sequence 4, Application US/08909983
; Patent No. 5747288
; GENERAL INFORMATION:
;   APPLICANT: Rubin, Gerry M.
;   APPLICANT: Therrien, Marc
;   APPLICANT: Chang, Henry C.
;   APPLICANT: Karim, Felix D.
;   APPLICANT: Wasserman, David A.
;   TITLE OF INVENTION: A No. 5747288el Protein Kinase Required for Ras
;   NUMBER OF SEQUENCES: 12
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
;   STREET: 268 BUSH STREET, SUITE 3200
;   CITY: SAN FRANCISCO
;   STATE: CALIFORNIA
;   COUNTRY: USA
;   ZIP: 94104
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patent In Release #1.0, Version #1.30
;   CURRENT APPLICATION NUMBER: US/08/909,983
;   FILING DATE: 12-JUN-1997
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/571,758
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: OSMAN, RICHARD A
;   REGISTRATION NUMBER: 36,627
;   REFERENCE/DOCKET NUMBER: B96-010
;   TELEPHONE: (415) 343-4341
;   TELEFAX: (415) 343-4342
;   INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1003 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: not relevant
;     TOPOLOGY: not relevant
;     MOLECULE TYPE: peptide
;
US-08-909-983-4

Query Match          9.7%; Score 66; DB 1; Length 1003;
Best Local Similarity 21.0%; Pred. No. 18;
Matches 26; Conservative 19; Mismatches 45; Indels 34; Gaps 4;

QY 3 SVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSA-----RRIGYGIKTTNM 50
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 602 SLTGSQVSTHSATSQVNSVSGSSSATYTSLLVNSGFFPRKLSNAGVDKRVFFSEYTD 661
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 51 KRLGVDPGCVLDPKRAVLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFR----- 105
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 662 HKSN-----DSDKTVSLS-----GSASTDSDRTPVRLDSTEDGDSGQWRQNSIS 705
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 106 -REW 108
   :||
Db 706 LKEW 709

Query Match          9.6%; Score 65.5; DB 2; Length 890;
Best Local Similarity 28.8%; Pred. No. 17;
Matches 32; Conservative 14; Mismatches 40; Indels 25; Gaps 6;

QY 8 DIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG--VLDPK 65
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 373 DVFVQAGDIINDNGRYVEKQKNHKSAINS-----GLRLPLTRNLAVQLCCGGAVIDNK 425
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 66 -----EAVLLAVSCDA---FAFGQEDTNNDRITVETWNTPDGAARQF 104
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
```


Db 426 NYETGILWNSGLDLSNSKFTFLFG-DDTHGNYQNVSYT---DGFSLSF 472

RESULT 9
US-08-786-455B-4
; Sequence 4, Application US/08786455B
; Patent No. 6193971
; GENERAL INFORMATION:
; APPLICANT: HOFMANN, Joachim
; APPLICANT: SCHMID, Karlheinerich
; TITLE OF INVENTION: DICTYOCALUS VIVIPARUS ANTIGEN FOR
; TITLE OF INVENTION: DIAGNOSING LUNGWORM INFESTATION AND FOR VACCINATION
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/087786,455B
; APPLICATION NUMBER: US/087786,455B
; FILING DATE: 21-JAN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 01 754.8
; FILING DATE: 19-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: GRANADOS, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 18748/327
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-786-455B-4

Query Match 9.5%; Score 64.5; DB 4; Length 14;
Best Local Similarity 86.7%; Pred. No. 0.062;
Matches 13; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 104 FRREWFQDGMVRRK 118
Db 1 FRREWFQDGMVRRK 14

RESULT 10
US-08-853-659A-47
; Sequence 47, Application US/08853659A
; Patent No. 5925522
; GENERAL INFORMATION:
; APPLICANT: Wong, K. K.; Saffer, J. D.
; TITLE OF INVENTION: A Salmonella Sequence, Methods Of Detection
; TITLE OF INVENTION: Of A
; TITLE OF INVENTION: Salmonella Sequence, And Methods Of Detection Of Salmonella
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Paul W. Zimmerman
; ADDRESSEE: Intellectual Property Services
; ADDRESSEE: Battelle Memorial Institute
; ADDRESSEE: PNNL P.O. Box 999

; STREET: Washington Way
; CITY: Richland
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 99352
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.50", 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (WordPerfect 5.1)
; CURRENT APPLICATION DATA: US/08/853,659A
; APPLICATION NUMBER: US/08/853,659A
; FILING DATE: Unknown
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: none
; FILING DATE: n/a
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 305 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-853-659A-47

Query Match 9.3%; Score 63.5; DB 2; Length 305;
Best Local Similarity 21.5%; Pred. No. 6.7;
Matches 35; Conservative 14; Mismatches 37; Indels 77; Gaps 8;
QY 6 PGDIQ-----TQPGTKIVFNAPY-----DKKHTYHKVI-----NSSARRI 41
Db 14 PGDVNEVSLSDGGVTWVKATQSATPGVWNYTWPCVTPDGDYTLNVKATDNAGVTETL 73
QY 42 GYGKTTNMKRLGVDPPCGVLDPKREAVLLAVSCDAFAFGQEDTNN----- 87
Db 74 HFTIDT-----LSTPVIVLDSAD-----DSGVHGDNMNTNSTQFTFALQHIDDDAV 119
QY 88 RITVE-----WNTPTDGAARQFRREWFQD 112
Db 120 RVTSVHEHGGVTTTFDTATKDAGGWTFTETGA-----WADGD 155

RESULT 11
US-08-553-501A-29
; Sequence 29, Application US/08553501A
; Patent No. 5856135
; GENERAL INFORMATION:
; APPLICANT: TSUCHIYA, Masayuki
; APPLICANT: SATO, Koh
; TITLE OF INVENTION: RESHAPED HUMAN ANTIBODY TO HUMAN
; TITLE OF INVENTION: INTERLEUKIN-6
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,501A
; FILING DATE: 20-FEB-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP94/00859
; FILING DATE: 30-MAY-1994
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 5-129787
FILING DATE: 31-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 53466/177/AAOK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-553-501A-29

Query Match 9.2%; Score 62.5; DB 2; Length 139;
Best Local Similarity 23.5%; Pred. No. 2.9;
Matches 20; Conservative 10; Mismatches 36; Indels 19; Gaps 4;
QY 43 YGKITTNNKRLGVDPCCGVLDPKKAVLLAVSCDAFAFGQEDTNNDRITVETWTPDGAAR 102
DB 16 YVLSQVTLKESGP-----GILQPSQTLSTLCSFSGFSL-----NTSGMTVGWIRQPSGKGL 66
QY 103 QFRRE-WFQGDGMVRKKNLP1EYNP 126
DB 67 EWLAIHWNDKY-----YNP 82

RESULT 12
US-09-205-231-29
Sequence 29, Application US/09205231
Patent No. 6121423
GENERAL INFORMATION:
APPLICANT: TSUCHIYA, Masayuki
APPLICANT: SATO, Koh
APPLICANT: HIRATA, Yuichi
TITLE OF INVENTION: RESHAPED HUMAN ANTIBODY TO HUMAN
TITLE OF INVENTION: INTERLEUKIN-6
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/553,501
FILING DATE: 20-FEB-1996
APPLICATION NUMBER: WO PCT/JP94/00859
FILING DATE: 30-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-129787
FILING DATE: 31-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: WEGNER, Harold C.
REGISTRATION NUMBER: 25,258
REFERENCE/DOCKET NUMBER: 53466/177/AAOK
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-205-231-29

Query Match 9.2%; Score 62.5; DB 3; Length 139;
Best Local Similarity 23.5%; Pred. No. 2.9;
Matches 20; Conservative 10; Mismatches 36; Indels 19; Gaps 4;
QY 43 YGKITTNNKRLGVDPCCGVLDPKKAVLLAVSCDAFAFGQEDTNNDRITVETWTPDGAAR 102
DB 16 YVLSQVTLKESGP-----GILQPSQTLSTLCSFSGFSL-----NTSGMTVGWIRQPSGKGL 66
QY 103 QFRRE-WFQGDGMVRKKNLP1EYNP 126
DB 67 EWLAIHWNDKY-----YNP 82

RESULT 13
US-08-867-941-19
Sequence 19, Application US/08867941
Patent No. 5977337
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Du, Run-Pan
APPLICANT: Wang, Quijun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1076 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-19

Query Match 9.1%; Score 62; DB 2; Length 1076;
Best Local Similarity 22.3%; Pred. No. 63;
Matches 23; Conservative 15; Mismatches 37; Indels 28; Gaps 4;
QY 8 DIQTQPGTKIVFNAPYDDKHT-----YHIKV-----INSARRIGYGI 45

Db 474 DVNKPFSVKEVDNNAYKEQHNLKAVFNKKMALGSTHHHNLQVGYDKFNSLSRVEYRL 533
 QY 46 KT-TNMKRLGVDPGCVLDPKAEVLLA-----VSCDAFAFGQE 82
 Db 534 ATHOSYQKLDYTPSPNPLDPKFKPILGSNNKPKICLDAYGYGHD 576

RESULT 14
 US-09-074-658-19
 ; Sequence 19, Application US/09074658
 ; Patent No. 6184371
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M
 ; APPLICANT: Run-Pan Du
 ; APPLICANT: Quijun Wang
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Klein, Michel H
 ; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
 ; NUMBER OF SEQUENCES: 78
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/074,658
 ; FILING DATE: 08-MAY-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-795
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1076 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-074-658-19

Query Match 9.1%; Score 62; DB 4; Length 1076;
 Best Local Similarity 22.3%; Pred. No. 63;
 Matches 23; Conservative 15; Mismatches 37; Indels 28; Gaps 4;
 QY 8 DIQTQPTGKIVFNAPYDDKHT-----YHIKV-----INSSARRIGYGI 45
 Db 474 DVNKPFSVKEVDNNAYKEQHNLKAVFNKKMALGSTHHHNLQVGYDKFNSLSRVEYRL 533
 QY 46 KT-TNMKRLGVDPGCVLDPKAEVLLA-----VSCDAFAFGQE 82
 Db 534 ATHOSYQKLDYTPSPNPLDPKFKPILGSNNKPKICLDAYGYGHD 576

RESULT 15
 PCT-US93-11404-2
 ; Sequence 2, Application PC/TUS9311404
 ; GENERAL INFORMATION:
 ; APPLICANT: THE BOARD OF TRUSTEES OF LELAND STANFORD
 ; APPLICANT: JUNIOR UNIVERSITY
 ; TITLE OF INVENTION: MUCOSAL VASCULAR ADDRESSIN, DNA AND
 ; TITLE OF INVENTION: EXPRESSION

NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Bertram I. Rowland
 STREET: 4 Embarcadero Center
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US93/11404
 FILING DATE: 23-Nov-1993
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Rowland, Bertram I.
 REGISTRATION NUMBER: 20,015
 REFERENCE/DOCKET NUMBER: FP-57452/BIR;STAN-144
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 405 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US93-11404-2

Query Match 9.0%; Score 61.5; DB 5; Length 405;
 Best Local Similarity 22.2%; Pred. No. 18;
 Matches 37; Conservative 15; Mismatches 52; Indels 63; Gaps 9;
 QY 2 QSVPPGDIQTOPGTKIVFNA-PYDDKHTYHIKVINSSARRIGYGIKTTNM----- 50
 Db 215 KEIPVLQSTSPKPPNTTSAEPY-----ILTSSSTAEAVSTGLNITTLPSAPPYPKLS 267
 QY 51 -KRLGVDPGCVLDPK-----EA-----VLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAA 101
 Db 268 PRTLSEGPC---RPKIHQDLEAGWELLCEASC-----GPGVTVTRWTLAPGDIA 313
 QY 102 RQFRE-----WFO-----GDGMVRRKNLPYENP 126
 Db 314 TYHKREAGAQAWLVLPPGPWVEGWFQCRQDPGGQVTNLYVPGQVTP 360

Search completed: September 20, 2002, 09:24:08
 Job time: 86 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:42 ; Search time 16.56 Seconds
(without alignments)
731.114 Million cell updates/sec

Title: US-09-863-063-2

Perfect score: 682

Sequence: 1 AQSVPFGDIQTQGTQKIVFN.....EFQGDGWMVRKRLPIEYNP 126

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 71:.*
1: pirl:.*
2: pirl2:.*
3: pirl3:.*
4: pirl4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	682	100.0	127	T16183	hypothetical prote
2	679	99.6	127	G88145	protein F58A6.8 [i
3	679	99.6	127	A88165	protein ZK1248.6 [i
4	679	99.6	127	G88686	protein msp-19 [im
5	679	99.6	127	C88688	protein msp-113 [i
6	679	99.6	127	H88688	protein msp-59 [im
7	679	99.6	127	H88689	protein msp-65 [im
8	679	99.6	127	C88689	protein msp-51 [im
9	679	99.6	127	H88792	protein K07F5.1 [i
10	679	99.6	127	H88146	protein C34F11.4 [i
11	679	99.6	127	E88134	protein msp-40 [im
12	679	99.6	127	F88138	protein MSP-31 [im
13	679	99.6	127	D88164	protein msp-142 [i
14	678	99.4	127	T24885	hypothetical prote
15	675	99.0	127	T21640	hypothetical prote
16	675	99.0	127	T16684	major sperm protei
17	675	99.0	127	A88139	hypothetical prote
18	675	99.0	133	T27902	hypothetical prote
19	673	98.7	127	F88146	protein C34F11.6 [i
20	671	98.4	127	C88164	protein K05F1.7 [i
21	670	98.2	127	A88683	protein C09B9.6 [i
22	668	97.9	127	F88801	protein C04G2.4 [i
23	665	97.5	127	T23486	hypothetical prote
24	639	93.7	130	T16687	major sperm protei
25	590	86.5	127	B45528	major sperm protei
26	585	85.8	127	A45528	major sperm protei
27	574	84.2	127	A45944	hypothetical prote
28	428	62.8	484	T26393	protein ZK1248.4 [i
29	418	61.3	77	F88165	

30	354.5	52.0	95	2	T34500	hypothetical prote
31	351	51.5	169	2	T31484	hypothetical prote
32	302	44.3	99	2	T31606	hypothetical prote
33	218	32.0	418	2	T19800	hypothetical prote
34	134	19.6	260	2	A57245	VAMP-binding prote
35	116.5	17.1	37	2	T29808	hypothetical prote
36	95	13.9	243	2	JG0186	vesicle-associated
37	90.5	13.3	245	2	T32774	hypothetical prote
38	84.5	12.4	185	2	T25098	hypothetical prote
39	82.5	12.1	319	2	T27116	hypothetical prote
40	79	11.6	352	2	AB2243	penicillin-binding
41	78.5	11.5	601	2	C81414	probable tenascin
42	77.5	11.4	4006	2	T09070	response regulator
43	77	11.3	612	2	F87690	hypothetical prote
44	76.5	11.2	221	2	T19317	hypothetical prote
45	75.5	11.1	1101	2	T21062	hypothetical prote

ALIGNMENTS

RESULT 1

T16183

hypothetical protein F26G1.7 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 26-May-2000

C:Accession: T16183

R:Chissoe, S.

submitted to the EMBL Data Library, July 1995/
A:Description: the sequence of *C. elegans* cosmid F26G1.

A:Reference number: Z18472

A:Accession: T16183

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-127 <CHI>

A:Cross-references: EMBL:U23519; NID:g746524; PID:g746531; PIDN:AAC46807.1; CESP:F26G

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:F26G1.7

C:Superfamily: *Caenorhabditis elegans* major sperm protein

Query Match 100.0%; Score 682; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 8.6e-65;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQSVPFGDIQTQGTQKIVFNAPYDDKHTYHKVINSARRIGYGIKTNNKRLGVDPPCG 60

Db 2 AQSVPFGDIQTQGTQKIVFNAPYDDKHTYHKVINSARRIGYGIKTNNKRLGVDPPCG 61

Qy 61 VLDPKAVLLAVSCDAFAFGQEDTNDRTVETWNTPDGAARQFRWFQGDGWMVRKML 120

Db 62 VLDPKAVLLAVSCDAFAFGQEDTNDRTVETWNTPDGAARQFRWFQGDGWMVRKML 121

Qy 121 PIEYNP 126

Db 122 PIEYNP 127

RESULT 2

G88145

protein F58A6.8 [imported] - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

C:Accession: G88145

R:Anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating bio

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: G88145

A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_II; PIDN:AA36204.1; PID:g1255857; GSPDB:GN00020; CESP:F58A6.8
C:Genetics:
A:Gene: F58A6.8
A:Map position: 2
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
Best Local Similarity 99.2%; Pred. No. 1.8e-64;
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIOTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
Db 2 AQSVPFGDIOTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKRL 120
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKRL 121
QY 121 PIEYNP 126
Db 122 PIEYNP 127

RESULT 3
A88165
protein ZK1248.6 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: A88165
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: A88165
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_II; PID:g862495; GSPDB:GN00020; CESP:ZK1248.6
A:Note: similar to C. elegans major sperm protein
C:Genetics:
A:Gene: ZK1248.6
A:Map position: 2
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
Best Local Similarity 99.2%; Pred. No. 1.8e-64;
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIOTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
Db 2 AQSVPFGDIOTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKRL 120
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKRL 121
QY 121 PIEYNP 126
Db 122 PIEYNP 127

RESULT 4
G88686
protein msp-19 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: G88686

R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88686
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_IV; PIDN:AA26926.1; PID:g3329619; GSPDB:GN00022
C:Genetics:
A:Gene: msp-19
A:Map position: 4
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
Best Local Similarity 99.2%; Pred. No. 1.8e-64;
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIOTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
Db 2 AQSVPFGDIOTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKRL 120
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKRL 121
QY 121 PIEYNP 126
Db 122 PIEYNP 127

RESULT 5
C88688
protein msp-113 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: C88688
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biological
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: C88688
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_IV; PIDN:AA42255.1; PID:g1825633; GSPDB:GN00022
C:Genetics:
A:Gene: msp-113
A:Map position: 4
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
Best Local Similarity 99.2%; Pred. No. 1.8e-64;
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIOTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
Db 2 AQSVPFGDIOTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKRL 120
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPTDGAARQFRREWFQDGMVRRKRL 121
QY 121 PIEYNP 126
Db 122 PIEYNP 127

```
RESULT 6
H88688
protein msp-59 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: H88688
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: H88688
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_IV; PIDN:AAB42253.1; PID:g1825631; GSPDB:GN00022
C:Genetics:
A:Gene: msp-59
A:Map position: 4
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
Best Local Similarity 99.2%; Pred. No. 1.8e-64;
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AOSVPPGDIQTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
|||||
Db 2 AOSVPPGDIQTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
|||||

Qy 61 VLDPKKAVLLAVSCDAFAFGQEDTNNDRITVETNTPDGAARQFRREWFQDGMVRRKNL 120
|||||
Db 62 VLDPKKAVLLAVSCDAFAFGQEDTNNDRITVETNTPDGAARQFRREWFQDGMVRRKNL 121
|||||

Qy 121 PIEYNP 126
|||||
Db 122 PIEYNP 127

RESULT 7
H88689
protein msp-65 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: H88689
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: H88689
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_IV; PIDN:AAB42256.1; PID:g1825634; GSPDB:GN00022
C:Genetics:
A:Gene: msp-65
A:Map position: 4
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
Best Local Similarity 99.2%; Pred. No. 1.8e-64;
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AOSVPPGDIQTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
|||||
Db 2 AOSVPPGDIQTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
|||||

Qy 61 VLDPKKAVLLAVSCDAFAFGQEDTNNDRITVETNTPDGAARQFRREWFQDGMVRRKNL 120
|||||
Db 62 VLDPKKAVLLAVSCDAFAFGQEDTNNDRITVETNTPDGAARQFRREWFQDGMVRRKNL 121
|||||

Qy 121 PIEYNP 126
|||||
Db 122 PIEYNP 127

RESULT 8
H88689
protein msp-51 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: H88689
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: H88689
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_IV; PIDN:AAB42254.1; PID:g1825632; GSPDB:GN00022
C:Genetics:
A:Gene: msp-51
A:Map position: 4
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
Best Local Similarity 99.2%; Pred. No. 1.8e-64;
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AOSVPPGDIQTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 60
|||||
Db 2 AOSVPPGDIQTQGTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPPCG 61
|||||

Qy 61 VLDPKKAVLLAVSCDAFAFGQEDTNNDRITVETNTPDGAARQFRREWFQDGMVRRKNL 120
|||||
Db 62 VLDPKKAVLLAVSCDAFAFGQEDTNNDRITVETNTPDGAARQFRREWFQDGMVRRKNL 121
|||||

Qy 121 PIEYNP 126
|||||
Db 122 PIEYNP 127

RESULT 9
H88792
protein K07F5.1 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: H88792
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: H88792
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_IV; PIDN:CAA94282.1; PID:g3878316; GSPDB:GN00022; CESP:K07
C:Genetics:
A:Gene: K07F5.1
A:Map position: 4
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
Best Local Similarity 99.2%; Pred. No. 1.8e-64;
```

Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60
 Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 61

QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 120
 Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 121

QY 121 PIEYNP 126
 Db 122 PIEYNP 127

RESULT 10
 H88146
 protein C34F11.4 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
 C:Accession: H88146
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; PMID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: H88146
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-127 <STO>
 A:Cross-references: GB:chr_II; PIDN:AA85761.1; PID:gl166627; GSPDB:GN000020; CESP:C34F11.4
 A:Note: similar to major sperm protein
 C:Genetics:
 A:Gene: C34F11.4
 A:Map position: 2
 C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
 Best Local Similarity 99.2%; Pred. No. 1.8e-64;
 Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60
 Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 61

QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 120
 Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 121

QY 121 PIEYNP 126
 Db 122 PIEYNP 127

RESULT 11
 H88134
 protein msp-40 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
 C:Accession: H88134
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
 A:Reference number: A75000; PMID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
 A:Accession: H88134
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-127 <STO>
 A:Cross-references: GB:chr_II; PIDN:AA93398.1; PID:gl203940; GSPDB:GN000020; CESP:C33F10.1

A:Note: C33F10.9
 C:Genetics:
 A:Gene: msp-40
 A:Map position: 2
 C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
 Best Local Similarity 99.2%; Pred. No. 1.8e-64;
 Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60
 Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 61

QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 120
 Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 121

QY 121 PIEYNP 126
 Db 122 PIEYNP 127

RESULT 12
 F88138
 protein MSP-31 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
 C:Accession: F88138
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998
 A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
 A:Reference number: A75000; PMID:99069613; PMID:9851916
 A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
 A:Accession: F88138
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-127 <STO>
 A:Cross-references: GB:chr_II; PIDN:AA83175.1; PID:gl109821; GSPDB:GN000020; CESP:R05F
 A:Note: R05F9.13
 C:Genetics:
 A:Gene: MSP-31
 A:Map position: 2
 C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
 Best Local Similarity 99.2%; Pred. No. 1.8e-64;
 Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60
 Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 61

QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 120
 Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 121

QY 121 PIEYNP 126
 Db 122 PIEYNP 127

RESULT 13
 D88164
 protein msp-142 [imported] - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
 C:Accession: D88164
 R:anonymous, The C. elegans Sequencing Consortium.
 Science 282, 2012-2018, 1998

A:Title: Genomé sequence of the nematode C. elegans: a platform for investigating biological
A:Reference number: A75000; MUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_elegans/
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: D88164
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <STO>
A:Cross-references: GB:chr_II; PID:9868174; GSPDB:GN00020; CESP:K05F1.2
A:Note: K05F1.2
C:Genetics:
A:Gene: msp-142
A:Map position: 2
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.6%; Score 679; DB 2; Length 127;
Best Local Similarity 99.2%; Pred. No. 1.8e-64;
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGYDPPCG 60
|||||
DB 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGYDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRFQDGMVRRKNL 120
|||||
DB 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRFQDGMVRRKNL 121
QY 121 PIEYNP 126
|||||
DB 122 PIEYNP 127

RESULT 14

T24885

hypothetical protein T13F2.11 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000

C:Accession: T24885

R:Swineburne, J.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19947

A:Accession: T24885

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-127 <WIL>

A:Cross-references: EMBL:Z81122; PIDN:CAB03362.1; GSPDB:GN00022; CESP:T13F2.11

A:Experimental source: clone T13F2

C:Genetics:

A:Gene: CESP:T13F2.11

A:Map position: 4

C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.4%; Score 678; DB 2; Length 127;
Best Local Similarity 98.4%; Pred. No. 2.3e-64;
Matches 124; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGYDPPCG 60
|||||
DB 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGYDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRFQDGMVRRKNL 120
|||||
DB 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRFQDGMVRRKNL 121
QY 121 PIEYNP 126
|||||
DB 122 PIEYNP 127

RESULT 15

T21640

hypothetical protein F32B6.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000
C:Accession: T21640; T24884
R:Basham, V.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19453
A:Accession: T21640
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-127 <WIL>
A:Cross-references: EMBL:Z81074; PIDN:CAB03037.1; GSPDB:GN00022; CESP:F32B6.6
A:Experimental source: clone F32B6
R:Swineburne, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19947
A:Accession: T24884
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-127 <WIL>
A:Cross-references: EMBL:Z81122; PIDN:CAB03361.1; GSPDB:GN00022; CESP:T13F2.10
A:Experimental source: clone T13F2
C:Genetics:
A:Gene: CESP:F32B6.6; CESP:T13F2.10
A:Map position: 4
C:Superfamily: Caenorhabditis elegans major sperm protein

Query Match 99.0%; Score 675; DB 2; Length 127;
Best Local Similarity 98.4%; Pred. No. 4.7e-64;
Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGYDPPCG 60
|||||
DB 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGYDPPCG 61
QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRFQDGMVRRKNL 120
|||||
DB 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRFQDGMVRRKNL 121
QY 121 PIEYNP 126
|||||
DB 122 PIEYNP 127

Search completed: September 20, 2002, 09:24:30
Job time: 108 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:42 ; Search time 11.89 Seconds
(without alignments)
410.317 Million cell updates/sec

Title: US-09-863-063-2
Perfect score: 682
Sequence: 1 AQSVPFGDIQTQPGTKIVFN.....EFQDGMVRRKNLPPIENP 126

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues 105224
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	679	99.6	126	1 MS31_CAEEL	P53017 caenorhabdi
2	675	99.0	126	1 MS33_CAEEL	P53019 caenorhabdi
3	665	97.5	126	1 MS38_CAEEL	P53020 caenorhabdi
4	663	97.2	126	1 MS10_CAEEL	P05634 caenorhabdi
5	658	96.5	126	1 MS56_CAEEL	P05635 caenorhabdi
6	639	93.7	190	1 MS32_CAEEL	P53018 caenorhabdi
7	590	86.5	126	1 MSP2_ASCSU	P27440 ascaris suu
8	590	86.5	126	1 MSP2_ONCVO	P13263 onchocerca
9	585	85.8	126	1 MSP1_ONCVO	P13262 onchocerca
10	574	84.2	126	1 MSP1_ASCSU	P27439 ascaris suu
11	436	63.9	125	1 MSP2_GLORO	P53022 globodera r
12	429	62.9	125	1 MSP1_GLORO	P53021 globodera r
13	425	62.3	125	1 MSP3_GLORO	P53023 globodera r
14	134	19.6	260	1 VF33_APLCA	Q16943 aplysia cal
15	74	10.9	131	1 YOD1_CAEEL	P34593 caenorhabdi
16	73	10.7	179	1 YNC5_CAEEL	P34538 caenorhabdi
17	71	10.7	244	1 SCS2_YEAST	P40075 saccharomyc
18	71	10.4	285	1 VRJ1_STRCO	P41108 streptomyc
19	69	10.1	863	1 MCM4_XENLA	P30664 xenopus lae
20	68.5	10.0	237	1 VT4_SFVKA	P25948 Shope fibro
21	68	10.0	221	1 SEGA_BPT4	P32286 bacterioph
22	67.5	9.9	996	1 VGNM_RCMV	P13561 red clover
23	67	9.8	328	1 RPOA_SHEPS	P74963 shewanella
24	67	9.8	436	1 SLS6_BRAOL	P07761 brassica ol
25	67	9.8	1630	1 PTP1_DROME	P35992 drosophila
26	67	9.8	4289	1 TENX_HUMAN	P22105 homo sapien
27	66	9.7	616	1 SPAN_STRPU	P98068 strongyloc
28	65	9.5	513	1 PPT1_YEAST	P53043 saccharomyc
29	65	9.5	1031	1 RAD2_YEAST	P07276 saccharomyc
30	64.5	9.5	333	1 LIPB_ARATH	Q23021 arabidopsis
31	64	9.4	553	1 PELW_FRWCH	Q05526 erwinia chr
32	64	9.4	2875	1 RRP1_STRWI	P28976 tomato spot
33	64	9.4	3067	1 CAIC_MOUSE	Q60847 mus musculu

34	63.5	9.3	333	1	TALL_KLUULA	P34214 kluyveromyc
35	63.5	9.3	1239	1	TOP2_CRIFA	P27570 crithidia f
36	63	9.2	895	1	YA54_METJA	Q58454 methanococc
37	63	9.2	1080	1	MI15_CAEEL	Q23356 caenorhabdi
38	62.5	9.2	337	1	G3P_USTMA	P09317 ustilago ma
39	62.5	9.2	503	1	IMDH_TRIFO	P50097 tritrichomo
40	62	9.1	322	1	YJ96_CAEEL	P49048 caenorhabdi
41	62	9.1	559	1	SUCP_ECOLI	P76041 escherichia
42	61.5	9.0	447	1	SUN_BACSU	P94464 bacillus su
43	61.5	9.0	818	1	PLD2_ORISA	P38844 oryza sativ
44	61	8.9	234	1	ENTD_SALTY	Q56064 salmonella
45	61	8.9	378	1	FENS_ORISA	P41345 oryza sativ

ALIGNMENTS

RESULT 1
MS31_CAEEL
ID MS31_CAEEL STANDARD; PRT; 126 AA.
AC P53017; 1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major sperm protein 31/40/142 (MSP).
GN (MSP-31 OR R05F9.13) AND (MSP-40 OR C33F10.9) AND (MSP-51 OR ZK354.5)
GN AND (MSP-59 OR ZK354.11) AND (MSP-65 OR ZK354.1) AND (MSP-113 OR
GN ZK354.4) AND (MSP-142 OR K05F1.2) AND C34F11.4 AND F58A6.8 AND K07F5.1
GN AND ZK1248.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Hallsworth K.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wohldmann P.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Bentley D.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Hemdriy C.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Latreille P.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A. (MSP-51; MSP-59; MSP-65 AND MSP-113).
RC STRAIN=BRISTOL N2;
RA Johnson D., Wamsley P., Bradshaw H.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
 CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
 CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
 CC -!- TISSUE SPECIFICITY: SPERM.
 CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
 CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC DR EMBL; U29244; AAC71087.1; -
 CC DR EMBL; U29377; AAA68711.1; -
 CC DR EMBL; U46753; AAA85761.1; -
 CC DR EMBL; U49830; AAA31477.1; -
 CC DR EMBL; U53339; AAA96204.1; -
 CC DR EMBL; U41533; AAA83175.1; -
 CC DR EMBL; Z70284; CAA94282.1; -
 CC DR EMBL; U88172; ABA42253.1; -
 CC DR EMBL; U88172; ABA42254.1; -
 CC DR EMBL; U88172; ABA42255.1; -
 CC DR EMBL; U88172; ABA42256.1; -
 CC DR HSSP; P27439; 3MSP.

CC DR WormPep; C33F10.9; CE02806.
 CC DR WormPep; C34F11.4; CE02806.
 CC DR WormPep; F58A6.8; CE02806.
 CC DR WormPep; K05F1.2; CE02806.
 CC DR WormPep; K07F5.1; CE02806.
 CC DR WormPep; R05F9.13; CE02806.
 CC DR WormPep; ZK1248.6; CE02806.
 CC DR WormPep; ZK354.1; CE09978.
 CC DR WormPep; ZK354.4; CE09978.
 CC DR WormPep; ZK354.5; CE09978.
 CC DR WormPep; ZK354.11; CE09978.
 CC DR InterPro; IPR000535; MSP_domain.
 CC Pfam; PF00635; MSP_domain; 1.

CC Cytoskeleton; Acetylation; Sperm; Multigene family.
 CC INIT_MET 0 0 BY SIMILARITY.
 CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 CC SEQUENCE 126 AA; 14078 MW; 0F069631D8559AB7 CRC64;

Query Match 99.6%; Score 679; DB 1; Length 126;
 Best Local Similarity 99.2%; Pred. No. 4.7e-66;
 Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHYYHKVINSARRIGYGIKTTNMKRLGVDPGPG 60
 DB 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHYYHKVINSARRIGYGIKTTNMKRLGVDPGPG 60
 QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNDRTITVETWNTPDGAARQFRFEGDGMVRRKNL 120
 DB 61 VLDPKAEVLLAVSCDAFAFGQEDTNDRTITVETWNTPDGAARQFRFEGDGMVRRKNL 120
 QY 121 PIEYNP 126
 DB 121 PIEYNP 126

RESULT 2
 MS33_CAEEL
 ID MS33_CAEEL STANDARD; PRT; 126 AA.
 AC P53019;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Major sperm protein 33 (MSP).
 GN MSP-33 OR R05F9.8.
 OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Hallsworth K.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
 CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
 CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
 CC -!- TISSUE SPECIFICITY: SPERM.
 CC -!- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
 CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC DR EMBL; U41533; AAA83165.1; -
 CC DR HSSP; P27439; 3MSP.
 CC DR WormPep; R05F9.8; CE04811.
 CC DR InterPro; IPR000535; MSP_domain.
 CC Pfam; PF00635; MSP_domain; 1.
 CC Cytoskeleton; Acetylation; Sperm; Multigene family.
 CC INIT_MET 0 0 BY SIMILARITY.
 CC MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
 CC SEQUENCE 126 AA; 14112 MW; 0F0C9C31D25F9AB7 CRC64;

Query Match 99.0%; Score 675; DB 1; Length 126;
 Best Local Similarity 98.4%; Pred. No. 1.3e-65;
 Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHYYHKVINSARRIGYGIKTTNMKRLGVDPGPG 60
 DB 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHYYHKVINSARRIGYGIKTTNMKRLGVDPGPG 60
 QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNDRTITVETWNTPDGAARQFRFEGDGMVRRKNL 120
 DB 61 VLDPKAEVLLAVSCDAFAFGQEDTNDRTITVETWNTPDGAARQFRFEGDGMVRRKNL 120
 QY 121 PIEYNP 126
 DB 121 PIEYNP 126

RESULT 3
 MS38_CAEEL
 ID MS38_CAEEL STANDARD; PRT; 126 AA.
 AC P53020;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Major sperm protein 38 (MSP).
 GN MSP-38 OR K08F4.8.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 CC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Hembry C.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
 CC SPERM CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
 CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
 CC -!- TISSUE SPECIFICITY: SPERM.

Query Match 96.5%; Score 658; DB 1; Length 126;
 Best Local Similarity 96.0%; Pred. No. 8.4e-64;
 Matches 121; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 AQSVPDGTOTQGTGKIVNAPYDDKHYYHUKVINSARRIGYGIKTTNNKRLGVDPGCG 60
 DB 1 AQSVPDGTOTQGTGKIVNAPYDDKHYYHUKVINSARRIGYGIKTTNNKRLGVDPGCG 60
 QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKRL 120
 DB 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKRL 120
 QY 121 PIEYNP 126
 DB 121 PIEYNP 126

RESULT 6
 MS32_CAEEL STANDARD; PRT; 190 AA.
 AC P33018;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Major sperm protein 32 (MSP).
 GN MSP-32 OR R05F9.3.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2.
 RA Hallsworth K.;
 RL Submitted (Nov-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
 CC Sperm crawling. Forms an extensive filament system that extends
 CC from sperm villipoda. Along the leading edge of the pseudopod.
 CC -1- TISSUE SPECIFICITY: SPERM.
 CC -1- MISCELLANEOUS: AROUND 30 MSP ISOFORMS MAY EXIST IN C.ELEGANS.
 CC -1- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U41533; AAA83168.1; .
 DR HSSP; P27439; 1MSP.
 DR WormPep; R05F9.3; CE04805.
 DR InterPro; IPR000535; MSP domain.
 DR Pfam; PF00635; MSP_domain; 1.
 KW Cytoskeleton; Sperm; Multigene family.
 SQ SEQUENCE 190 AA; 21304 MW; 939C5BBFD79C54BA CRC64;

Query Match 93.7%; Score 639; DB 1; Length 190;
 Best Local Similarity 98.3%; Pred. No. 1.5e-61;
 Matches 118; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 7 GDTQTPGTGKIVNAPYDDKHYYHUKVINSARRIGYGIKTTNNKRLGVDPGCGVLPKE 66
 DB 71 GMITQTPGTGKIVNAPYDDKHYYHUKVINSARRIGYGIKTTNNKRLGVDPGCGVLPKE 130
 QY 67 AVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKRLPIEYNP 126
 DB 131 AVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKRLPIEYNP 190

RESULT 7
 MSP2_ASCSU STANDARD; PRT; 126 AA.
 ID MSP2_ASCSU
 AC P27440;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Major sperm protein, isoform beta (beta-MSP).
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 OC Ascarididae; Ascaris.
 OX NCBI_TaxID=6253;
 RN [1]
 RP SEQUENCE.
 RC TISSUE-Sperm;
 RX MEDLINE=92407055; PubMed=1527183;
 RA King K.L.; Stewart M.; Roberts T.M.; Seavy M.;
 RT "Structure and macromolecular assembly of two isoforms of the major
 RT sperm protein (MSP) from the amoeboid sperm of the nematode, Ascaris
 RT suum.";
 RL J. Cell Sci. 101:847-857(1992).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS), AND REVISION TO 113.
 RX MEDLINE=98162552; PubMed=9501910;
 RA Bullock T.L.; McCoy A.J.; Kent H.M.; Roberts T.M.; Stewart M.;
 RT "Structural basis for amoeboid motility in nematode sperm.";
 RL Nat. Struct. Biol. 5:184-189(1998).
 CC -1- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
 CC Sperm crawling. Forms an extensive filament system that extends
 CC from sperm villipoda. Along the leading edge of the pseudopod.
 CC -1- SUBUNIT: FORMS FILAMENTS 10 NM WIDE, WITH A CHARACTERISTIC
 CC SUBSTRUCTURE REPEATING AXIALLY AT 9 NM.
 CC -1- TISSUE SPECIFICITY: SPERM.
 CC -1- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.

DR PDB; 2MSP; 15-APR-98.
 DR InterPro; IPR000535; MSP_domain; 1.
 DR Pfam; PF00635; MSP_domain; 1.
 KW Cytoskeleton; Acetylation; Sperm; Multigene family; 3D-structure.
 FT INIT_MET 0 BY SIMILARITY.
 FT MOD_RES 1 1 ACETYLATION.
 FT CONFLICT 113 113 G > D (IN REF. 1).
 SQ SEQUENCE 126 AA; 14117 MW; 177E4AFFB98850C3 CRC64;

Query Match 86.5%; Score 590; DB 1; Length 126;
 Best Local Similarity 84.0%; Pred. No. 1.7e-56;
 Matches 105; Conservative 9; Mismatches 11; Indels 0; Gaps 0;

QY 1 AQSVPDGTOTQGTGKIVNAPYDDKHYYHUKVINSARRIGYGIKTTNNKRLGVDPGCG 60
 DB 1 AQSVPDGTOTQGTGKIVNAPYDDKHYYHUKVINSARRIGYGIKTTNNKRLGVDPGCG 60
 QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKRL 120
 DB 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVEWNTPDGAARQFRREWFQDGMVRRKRL 120
 QY 121 PIEYNP 125
 DB 121 PIEYNP 125

RESULT 8
 MSP2_ONCVO STANDARD; PRT; 126 AA.
 ID MSP2_ONCVO
 AC P13263;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE Major sperm protein 2 (MSP2).
 OS Onchocerca volvulus.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 OC Onchocercidae; Onchocerca.

```
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04662; -; NOT_ANNOTATED_CDS.
DR PIR; A45528; A45528.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
DR Cytoskeleton; Acetylation; Sperm; Multigene family.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
SQ SEQUENCE 126 AA; 14211 MW; D308E525B511EC81 CRC64;
-----
Query Match 85.8%; Score 585; DB 1; Length 126;
Best Local Similarity 84.0%; Pred. No. 5.9e-56;
Matches 105; Conservative 8; Mismatches 12; Indels 0; Gaps 0;
-----
QY 1 AQSVPGGDIOTPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTNMKRLGVDPGCG 60
DB 1 AQSVPGGDIHTPGSKIVFNAPYDDKHTYHINTNAGGRIGWAIKTNNMKRLGVDPGCG 60
QY 61 VLDPKKAVLLAVSCDAFAFGQEDTNNDRITVWNTPDGAARQFRREWFQDGMVRRKNL 120
DB 61 VLDPKENVLMAVSCDTFDATREDINNDRIETWNTPDGAARQFRREWFQDGMVRRKNL 120
QY 121 PIEYN 125
DB 121 PIEYN 125
-----
RESULT 10
MSPL_ASCSU STANDARD; PRT; 126 AA.
AC P27439; P27441;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Major sperm protein, isoform alpha (Alpha-MSP).
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE.
RC TISSUE-Sperm;
RX MEDLINE=92407055; PubMed=1527183;
RA King K.L., Stewart M., Roberts T.M., Seavy M.;
RT "Structure and macromolecular assembly of two isoforms of the major
RT sperm protein (MSP) from the amoeboid sperm of the nematode, Ascaris
RT suum.";
RL J. Cell Sci. 101:847-857(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87031211; PubMed=3770294;
RA Bennett K.L., Ward S.;
RT "Neither a germ line-specific nor several somatically expressed genes
RT are lost or rearranged during embryonic chromatin diminution in the
RT nematode Ascaris lumbricoides var. suum.";
RL Dev. Biol. 118:141-147(1986).
RN [3]
RP SEQUENCE FROM N.A.
RA Bullock T.L., Parathasathy G., King K.L., Kent M.L., Roberts T.M.,
RA Stewart M.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=97070381; PubMed=8913307;
-----
CC CC
```

```

RA Bullock T.L., Roberts T.M., Stewart M.;
RT "2.5-A resolution crystal structure of the motile major sperm protein
RL (MSP) of Ascaris sum.";
RN J. Mol. Biol. 263:284-296(1996).
RP STRUCTURE BY NMR.
RX MEDLINE-99096891; PubMed-9878374;
RA Haaf A., Leclaire L. III, Roberts G., Kent H.M., Roberts T.M.,
RA Stewart M., Neuhaus D.;
RT "Solution structure of the motile major sperm protein (MSP) of
RT Ascaris sum - evidence for two manganese binding sites and the
RT possible role of divalent cations in filament formation.";
RN J. Mol. Biol. 284:1611-1624(1998).
RP
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD.
CC -!- SUBUNIT: FORMS FILAMENTS 10 NM WIDE, WITH A CHARACTERISTIC
CC SUBSTRUCTURE REPEATING AXIALLY AT 9 NM.
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15680; AAA29375.1; -.
DR EMBL; X94249; CAA63933.1; -.
DR PIR; A45944; A45944.
DR PDB; 1MSP; 07-DEC-96.
DR PDB; 3MSP; 20-APR-99.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
DR Cytoskeleton; Acetylation; Sperm; Multigene family.
KW INIT_MET 0
FT MOD_RES 1 1 ACETYLATION.
FT CONFLICT 113 113 G->D (IN REF. 1).
FT SEQUENCE 126 AA; 14259 MW; 477DCE6F4CFDD8F CRC64;
SQ
Query Match 84.2%; Score 574; DB 1; Length 126;
Best Local Similarity 82.4%; Pred. No. 8,9e-55;
Matches 103; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
Db 1 AQSVPVPGDINTQPSQKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60

QY 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETNTPDGAARQFRFQDGMVRRKNL 120
Db 61 VLDPKAVLLAVSCDTFNATDNLNDRITVETNTPDGAARQFRFQDGMVRRKNL 120

QY 121 PIEYN 125
Db 121 PIEYN 125

RESULT 11
MSP2_GLORO STANDARD; PRT; 125 AA.
AC P53022;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 2.
GN MSP-2.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=ROI;
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD (BY
CC SIMILARITY).
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L24500; AAA29147.1; -.
DR HSP; P27439; IMSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
DR Cytoskeleton; Acetylation; Sperm; Multigene family.
KW INIT_MET 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT SEQUENCE 125 AA; 13801 MW; 1B04AD5756511B13 CRC64;
SQ
Query Match 63.9%; Score 436; DB 1; Length 125;
Best Local Similarity 63.9%; Pred. No. 6e-40;
Matches 78; Conservative 16; Mismatches 28; Indels 0; Gaps 0;

QY 4 VPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCGVLD 63
Db 3 LPPGDIATMPNOKVFNAPDNKATYYVIVNPGTNRIGFAFKTKPKRINNPNNGVLG 62

QY 64 PKEAVLLAVSCDAFAFGQEDTNNDRITVETNTPDGAARQFRFQDGMVRRKNLPIE 123
Db 63 PKEVNVVAISDAFDPSSDTGDRVTVMCNTPDPAFAAFKLEWFGDGMVRRKNLPIE 122

QY 124 YN 125
Db 123 YN 124

RESULT 12
MSP1_GLORO STANDARD; PRT; 125 AA.
AC P53021;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Major sperm protein 1.
GN MSP-1.
OS Globodera rostochiensis (Golden nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
OX NCBI_TaxID=31243;
RN [1]
RP SEQUENCE FROM N.A.
RA Novitski C.E., Brown S., Chen R., Corner A.S., Atkinson H.J.,
RA McPherson M.J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CENTRAL COMPONENT IN MOLECULAR INTERACTIONS UNDERLYING
CC Sperm CRAWLING. FORMS AN EXTENSIVE FILAMENT SYSTEM THAT EXTENDS
CC FROM SPERM VILLIPODA, ALONG THE LEADING EDGE OF THE PSEUDOPOD (BY
CC SIMILARITY).
CC -!- TISSUE SPECIFICITY: SPERM.
CC -!- SIMILARITY: BELONGS TO THE NEMATODE MSP FAMILY.

```


Query Match 62.38; Score 425; DB 1; Length 125;

4 VDDCDIOTOPCTKTVE NARVYDNDKHTYHTKVINSABRIGYGIKTTNMKRLGVDP PCGVLD 63

[illegible]

THE

```
QY      PREAVLLVASCDAFAFGQEDINNDRIIVEWINIFEGARRQAKRENTQOCCORVRRARETLE
64      |||.:.:||| ||| :||| ||| | :||| ||| ||| |||
```

Db 63 PKESVNVAISCDAFDPSSSEDSKGDVRTVEWCNTPDPAAAAFKLEWFQGDGMVRRKKNLPTE 12

OY 124 YN 125

123 VN 124

RESULT 14

VP33_APLCA
ID VP33 APLCA
STANDARD: PRT: 260 AA.

AC Q16943;

DT 01-NOV-1997 (Rel. 35, Last sequence update)

BT 01-NOV-1997 (REL: 33, LAST annotation update)
 DE Vesicle-associated membrane protein/synaptobrevin binding protein

DE (VAP-33).

OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anaspidea

UC Aplysiidae; Aplysia.
 OX NCBI TaxID=6500:

RN [1] CONFERENCE FROM N A

RX MEDLINE=95397152; PubMed=7667638;

"A VAMP-binding protein from Aplysia required for neurotransmitter

release." ;
250.1590-1593/1095)

CC - ! - FUNCTION: REQUIRED FOR NEUROTRANSMITTER RELEASE. INTERACTS WITH

-!- SUBCELLULAR LOCATION: Membrane-bound.

-I- TISSUE SPECIFICITY: DETECTED ONLY IN THE CENTRAL NERVOUS SYSTEM AND THE GILL OF APLYSIA

CC-----through a collabor

between the Swiss Institute of Bioinformatics and the EMBL outsta

the European Bioinformatics Institute: there are no restrictions on the use of non-profit institutions as long as its content is in

cc modified and this statement is not removed. Usage by and for commercial purposes is not permitted. For more information, see <http://www.ibm.com/copying/>.

CC or send an email to license@isb-sib.ch).

EMBL: U36779: AAC46883.1: -

```
DR InterPro; IPR000535; MSP_domain.
```

KW Synapse; Transmembrane; Synaptosome.

FT	TRANSMEM	239	235	ORIENTAT.
SO	SEQUENCE	260 AA:	28328 MW:	0F01CABE63C732FE CRC64;

Query Match 19.6%; Score 134; DB 1; Length 260;

Best Local Similarity	25.36	22.64	22.64
Matches	29	Conservative	18
Mismatches	18	Mismatches	50
Indels	50	Indels	0
Gaps	0	Gaps	0

13 00CTKTYENAPVDDKHVTYVHIKVINSSARPTGYGIKTTNMKRLGVDPCCGVLDPKEAVLLA 71

70

DD IT EFAGEENT KOT E TDA V ANDERSON E BARKER FIVE

```
QY 72 VSCDAFAFGQEDTNNDRITVNTPTDGAARQFRREW 108
| | | | | | | | | | | | | | | | | | | | |
Db 71 VMLQPNYDPNEKNKHKFWQSWYAPDRHVESQELLW 107
| | | | | | | | | | | | | | | | | | | | |

RESULT 15
YODL_CAEEL
ID YODL_CAEEL STANDARD; PRT; 131 AA.
AC P34593;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 15.2 kDa protein ZC262.1 in chromosome III.
GN ZC262.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisner N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohlman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L23647; AAA28175.1; -
DR PIR; S44880; S44880.
DR WormPep; ZC262.1; CE00349.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
DR KW Hypothetical protein.
SQ SEQUENCE 131 AA; 15167 MW; 59FA944CD6EE48FA CRC64;

Query Match 10.9%; Score 74; DB 1; Length 131;
Best Local Similarity 25.7%; Pred. No. 0.49;
Matches 26; Conservative 18; Mismatches 43; Indels 14; Gaps 5;

QY 16 KIVENAPYDDKH-TYHLKVINSSARRIGYIKTNMKRLGVDPGVLDPKEAVLLAV-- 72
| | | | | | | | | | | | | | | | | | | | |
Db 22 EVEFKCTEDRKPISNVLKLNHTAVTVSYKVRCTSDIFRVQPLGLGVKPSVTSIVIW 81
| | | | | | | | | | | | | | | | | | | | |

QY 73 ----SCDAFA----FGQEDTNNDRITVE--WTNTP-DGAAR 102
| | | | | | | | | | | | | | | | | | | | |
Db 82 QNQDKDAISKNIHYFAFYHTNSDGTARELWANSKEGVGR 122
| | | | | | | | | | | | | | | | | | | | |
```

Search completed: September 20, 2002, 09:23:35
Job time: 53 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 09:22:42 ; Search time 24.67 Seconds
(without alignments)
883.557 Million cell updates/sec

Title: US-09-863-063-2
Perfect score: 682
Sequence: 1 AOSVPPGDIQTQPGTKIVFN.....EWFQGDGMVRRKNLPPIENP 126

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	682	100.0	127	5 Q19832	Q19832 caenorhabdi
2	679	99.6	130	5 Q9N2M1	Q9n2m1 caenorhabdi
3	678	99.4	127	5 Q94053	Q94053 caenorhabdi
4	676	99.1	127	5 Q95XU7	Q95xu7 caenorhabdi
5	675	99.0	127	5 Q9TVW5	Q9tvw5 caenorhabdi
6	675	99.0	127	5 Q23519	Q23519 caenorhabdi
7	675	99.0	133	5 Q23524	Q23524 caenorhabdi
8	673	98.7	127	5 Q18461	Q18461 caenorhabdi
9	671	98.4	127	5 Q21244	Q21244 caenorhabdi
10	670	98.2	127	5 Q17856	Q17856 caenorhabdi
11	668	97.9	127	5 Q27280	Q27280 caenorhabdi
12	498	73.0	107	5 Q9GNV8	Q9gnv8 onchocerca
13	496	72.7	107	5 Q9GNV9	Q9gnv9 onchocerca
14	494	72.4	107	5 Q9GNW0	Q9gnw0 onchocerca
15	493	72.3	107	5 Q9GN10	Q9gn10 mansoniella
16	492	72.1	107	5 Q9GNW2	Q9gnw2 onchocerca

17	490	71.8	107	5 Q9GNW6	Q9gnw6 mansoniella
18	488	71.6	107	5 Q9GNX0	Q9gnx0 mansoniella
19	488	71.6	107	5 Q9GNW7	Q9gnw7 mansoniella
20	487	71.4	107	5 Q9GNX1	Q9gnx1 mansoniella
21	485	71.1	107	5 Q9GNW1	Q9gnw1 onchocerca
22	485	71.1	107	5 Q9GNW9	Q9gnw9 mansoniella
23	484	71.0	107	5 Q9GNW8	Q9gnw8 mansoniella
24	484	71.0	107	5 Q9GNW4	Q9gnw4 mansoniella
25	483	70.8	107	5 Q9GNW5	Q9gnw5 mansoniella
26	428	62.8	442	5 Q9NAP2	Q9nap2 caenorhabdi
27	428	62.8	484	5 Q9NAM2	Q9nam2 caenorhabdi
28	418	61.3	77	5 Q95PJ7	Q95pj7 caenorhabdi
29	357	52.3	90	5 Q26098	Q26098 pratylenchu
30	354.5	52.0	95	5 Q23428	Q23428 caenorhabdi
31	351	51.5	167	5 Q9U2V8	Q9u2v8 caenorhabdi
32	351	51.5	169	5 Q9U2W1	Q9u2w1 caenorhabdi
33	339	49.7	88	5 Q27405	Q27405 pratylenchu
34	329	48.2	85	5 Q26097	Q26097 pratylenchu
35	309	45.3	84	5 Q26111	Q26111 pratylenchu
36	305	44.7	83	5 Q26096	Q26096 pratylenchu
37	302	44.3	99	5 Q9NAF5	Q9naf5 caenorhabdi
38	293	43.0	83	5 Q26112	Q26112 pratylenchu
39	279	40.9	141	5 Q26316	Q26316 dictyocaulu
40	218	32.0	418	5 Q18503	Q18503 caenorhabdi
41	116.5	17.1	37	5 Q18145	Q18145 caenorhabdi
42	114	16.7	242	4 Q9UBZ2	Q9ubz2 homo sapien
43	114	16.7	242	4 Q75453	Q75453 homo sapien
44	114	16.7	242	4 Q9FOL0	Q9fol0 homo sapien
45	114	16.7	242	11 Q9Z270	Q9z270 rattus norv

ALIGNMENTS

RESULT 1
Q19832 Q19832 PRELIMINARY; PRT; 127 AA.
AC Q19832;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 14.2 KDA PROTEIN.
GN F26G1.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Chisoe S.;
RT "The sequence of C. elegans cosmid F26G1.";
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U23519; RAK31504.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
DR Hypothetical protein.
SQ SEQUENCE 127 AA; 14237 MW; 71671F31BEA5B147 CRC64;

```
Query Match          100.0%; Score 682; DB 5; Length 127;
Best Local Similarity 100.0%; Pred. No. 7.3e-66;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQGTGKIVFNAPYDDKHYYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60
Db 2 AQSVPVPGDIQTQGTGKIVFNAPYDDKHYYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 61
QY 61 VLDPKKAVLLAVSCDAFAFGQEDTNDRIIVETWNTPDGAARQFRREWFQDGMVRKKNL 120
Db 62 VLDPKKAVLLAVSCDAFAFGQEDTNDRIIVETWNTPDGAARQFRREWFQDGMVRKKNL 121
QY 121 PIEYNP 126
Db 122 PIEYNP 127

RESULT 2
Q9N2M1 ID Q9N2M1 PRELIMINARY; PRT; 130 AA.
AC Q9N2M1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Y59E9AR.1 PROTEIN (Y59E9AR.7 PROTEIN).
GN Y59E9AR.1 OR Y59E9AR.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024839; AAF60826.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
SQ SEQUENCE 130 AA; 14504 MW; 063BB2E75DE45198 CRC64;

Query Match          99.6%; Score 679; DB 5; Length 130;
Best Local Similarity 99.2%; Pred. No. 1.6e-65;
Matches 125; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQGTGKIVFNAPYDDKHYYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60
Db 5 AQSVPVPGDIQTQGTGKIVFNAPYDDKHYYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 64
QY 61 VLDPKKAVLLAVSCDAFAFGQEDTNDRIIVETWNTPDGAARQFRREWFQDGMVRKKNL 120
Db 65 VLDPKKAVLLAVSCDAFAFGQEDTNDRIIVETWNTPDGAARQFRREWFQDGMVRKKNL 124
QY 121 PIEYNP 126
Db 125 PIEYNP 130

RESULT 3
Q94053 ID Q94053 PRELIMINARY; PRT; 127 AA.
```

```
Q94053;
AC 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE T13F2.11 PROTEIN.
GN T13F2.11.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: A platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81122; CAB03362.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
SQ SEQUENCE 127 AA; 14223 MW; 69F69622069E2147 CRC64;

Query Match          99.4%; Score 678; DB 5; Length 127;
Best Local Similarity 98.4%; Pred. No. 2e-65;
Matches 124; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AQSVPVPGDIQTQGTGKIVFNAPYDDKHYYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 60
Db 2 AQSVPVPGDIQTQGTGKIVFNAPYDDKHYYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCG 61
QY 61 VLDPKKAVLLAVSCDAFAFGQEDTNDRIIVETWNTPDGAARQFRREWFQDGMVRKKNL 120
Db 62 VLDPKKAVLLAVSCDAFAFGQEDTNDRIIVETWNTPDGAARQFRREWFQDGMVRKKNL 121
QY 121 PIEYNP 126
Db 122 PIEYNP 127

RESULT 4
Q95XU7 ID Q95XU7 PRELIMINARY; PRT; 127 AA.
AC Q95XU7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE HYPOTHETICAL 14.2 KDA PROTEIN.
GN Y59H11AM.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Layman D.;
RT "The sequence of C. elegans cosmid Y59H11AM.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
```

RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC024841; AAK68539.1; -
KW Hypothetical protein.
SQ SEQUENCE 127 AA; 14195 MW; 69EA78C3A24B4347 CRC64;

Query Match 99.1%; Score 676; DB 5; Length 127;
Best Local Similarity 98.4%; Pred. No. 3.2e-65;
Matches 124; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
Db 2 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61

Qy 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 120
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 121

Qy 121 PIEYNP 126
Db 122 PIEYNP 127

RESULT 5
ID Q9TWW5 PRELIMINARY; PRT; 127 AA.
AC Q9TWW5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE T13F2.10 PROTEIN.
GN T13F2.10 OR F32B6.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Swaburne J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Basham V.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z81122; CAB03361.1; -
DR EMBL; Z81074; CAB03037.1; -
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
SQ SEQUENCE 127 AA; 14181 MW; 69F6962335A5B147 CRC64;

Query Match 99.0%; Score 675; DB 5; Length 127;
Best Local Similarity 98.4%; Pred. No. 4.1e-65;
Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
Db 2 AQSVPGGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61

Qy 61 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 120
Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 121

Db 62 VLDPKAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMARRKNL 121
Qy 121 PIEYNP 126
Db 122 PIEYNP 127

RESULT 6
ID Q23519 PRELIMINARY; PRT; 127 AA.
AC Q23519;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 14.2 KDA PROTEIN.
GN ZK546.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Hallsworth K.;
RT "The sequence of C. elegans cosmid ZK546.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U29380; AAA68739.1; -
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.
SQ SEQUENCE 127 AA; 14239 MW; 69F6962DCEACC152 CRC64;

Query Match 99.0%; Score 675; DB 5; Length 127;
Best Local Similarity 99.2%; Pred. No. 4.1e-65;
Matches 124; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 QSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61
Db 3 QSVPPGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 62

Qy 62 LDPKEAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNLP 121
Db 63 LDPKEAVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNLP 122

Qy 122 IEYNP 126
Db 123 IEYNP 127

RESULT 7
ID Q23524 PRELIMINARY; PRT; 133 AA.
AC Q23524;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 14.9 KDA PROTEIN.

GN ZK546.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Hallsworth K.;
RT "The sequence of C. elegans cosmid ZK546.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U29380; AAA68736.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.
SQ SEQUENCE 133 AA; 14899 MW; B283C4CDFDF9B76A CRC64;

Query Match 99.0%; Score 675; DB 5; Length 133;
Best Local Similarity 98.4%; Pred. No. 4.4e-65;
Matches 124; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AOSVPPGDIQTQPTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
|||||
Db 8 AOSVPPGDIQTQPTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 67
|||||
QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 120
|||||
Db 68 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 127
|||||
QY 121 PIEYNP 126
Db 128 PIEYNP 133
|||||

RESULT 8
Q18461 ID Q18461 PRELIMINARY; PRT; 127 AA.
AC Q18461;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 14.2 KDA PROTEIN.
GN C34F11.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;
RA Bentley D.;
RT "The sequence of C. elegans cosmid C34F11.";
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U46753; AAA85759.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.
SQ SEQUENCE 127 AA; 14221 MW; 53D359DB8EA5AB74 CRC64;

Query Match 98.7%; Score 673; DB 5; Length 127;
Best Local Similarity 98.4%; Pred. No. 6.8e-65;
Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AOSVPPGDIQTQPTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 60
|||||
Db 2 AOSVPPGDIQTQPTGKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCG 61
|||||
QY 61 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 120
|||||
Db 62 VLDPKAEVLLAVSCDAFAFGQEDTNNDRITVETWNTPDGAARQFRREWFQDGMVRRKNL 121
|||||
QY 121 PIEYNP 126
Db 122 PIEYNP 127
|||||

RESULT 9
Q21244 ID Q21244 PRELIMINARY; PRT; 127 AA.
AC Q21244;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE HYPOTHETICAL 14.2 KDA PROTEIN.
GN K05F1.7
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wohldmann P.;
RT "The sequence of C. elegans cosmid K05F1.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U29377; AAA68714.1; -.
DR HSSP; P27439; 3MSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
KW Hypothetical protein.

SQ SEQUENCE 127 AA; 14228 MW; 7D5D628EF5511AA7 CRC64;

Query Match 98.4%; Score 671; DB 5; Length 127;
Best Local Similarity 98.4%; Pred. No. 1.1e-64;
Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMRRLGVDPGCG 60
|||||
Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMRRLGVDPGCG 61
|||||

QY 61 VLDPKKAVLLAVSCDAFAFGQEDTNDRTITVETWTPDGAARQFRFQDGMVRRKNL 120
|||||
Db 62 VLDPKKAVLLAVSCDAFAFGQEDTNDRTITVETWTPDGAARQFRFQDGMVRRKNL 121
|||||

QY 121 PIEYNP 126
|||||

Db 122 PIEYNP 127
|||||

RESULT 10

ID Q17856 PRELIMINARY; PRT; 127 AA.
AC Q17856;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE C09B9.6 PROTEIN (HYPOTHETICAL PROTEIN R13H9.2).

GN C09B9.6 OR R13H9.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RAF Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,

RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans";

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Bradshaw H.;

RT "The sequence of C. elegans cosmid C09B9.6";

RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waterston R.;

RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RX MEDLINE=99069613; PubMed=9851916;

RA None;

RT "Genome sequence of the nematode C. elegans: a platform for

RT investigating biology. The C. elegans Sequencing Consortium.";

RL Science 282:2012-2018(1998).

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RT "The sequence of C. elegans cosmid R13H9.6";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE FROM N.A.

RC STRAIN=BRISTOL N2;

RA Waterston R.;

RT "Direct Submission";

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; U50069; AAB37555.1; -;

DR EMBL; AC006681; AAK85492.1; -;

DR HSSP; P27439; 3MSP

DR InterPro; IPR000535; MSP_domain.

DR Pfam; PF00635; MSP_domain; 1.

SQ SEQUENCE 127 AA; 14251 MW; 28E69731A2D5AD2B CRC64;

Query Match 98.2%; Score 670; DB 5; Length 127;

Best Local Similarity 98.4%; Pred. No. 1.4e-64;

Matches 124; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMRRLGVDPGCG 60
|||||

Db 2 AQSVPFGDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMRRLGVDPGCG 61
|||||

QY 61 VLDPKKAVLLAVSCDAFAFGQEDTNDRTITVETWTPDGAARQFRFQDGMVRRKNL 120
|||||

Db 62 VLDPKKAVLLAVSCDAFAFGQEDTNDRTITVETWTPDGAARQFRFQDGMVRRKNL 121
|||||

QY 121 PIEYNP 126
|||||

Db 122 PIEYNP 127
|||||

RESULT 11

Q27280

ID Q27280 PRELIMINARY; PRT; 127 AA.

AC Q27280;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE C04G2.4 PROTEIN.

GN C04G2.4 OR ZK1251.6 OR K07F5.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Hembrly C.;

RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RAF Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,

RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans";

RL Nature 368:32-38(1994).

RN [3]

RP SEQUENCE FROM N.A.

RA McMurray A.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Hembrly C.;

RU Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases
 DR EMBL; Z70718; CAA94674.1; -
 DR EMBL; Z68222; CAA92502.1; -
 DR EMBL; Z70284; CAA94278.1; -
 DR EMBL; Z70284; CAA94283.1; -
 DR HSSP; Z72749; 3MSP.
 DR InterPro; IPR000535; MSP_domain.
 DR Pfam; PF00635; MSP_domain; 1W.
 SQ SEQUENCE 127 AA; 14236 MW; ESB96631BEBF1419 CRC64;

[illegible]

RESULT	12
Q9GNV8	
ID	PRELIMINARY; PRT; 107 AA.
AC	Q9GNV8
DT	01-VAR-2001 (TReMBLrel. 16, Created)
DT	01-VAR-2001 (TReMBLrel. 16, Last sequence update)
DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE	MAJOR SPERM PROTEIN (FRAGMENT).
DE	MSP.
OS	Onchocerca volvulus.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC	Onchocercidae; Onchocerca.
OX	NCBI_TaxID=6282;
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=OVNODERA2C5 ;
RC	MEDLINE=11428336;
RX	Morales Hojas R., Post R.J.;
RA	"Regional genetic variation in the major sperm protein genes of
RT	Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea).";
RT	Int. J. Parasitol. 30:1459-1465(2000).
RL	EMBL; AJ404208; CAC20742.1; -
DR	HSSP; P27439; INSP.
DR	InterPro; IPR000535; MSP_domain.
DR	Pfam; PF00635; MSP_domain; 1.
FT	NON_TER 1
FT	NON_TER 1
FT	NON_TER 107
SQ	SEQUENCE 107 AA; 111987 MW; OC4325F7F6E5DF07 CRC64:

	Query Match	73.0%	Score 498;	DB 5;	Length 107;
	Best Local Similarity	82.2%	Pred. No. 4e-46;		
	Matches	88;	Conservative	8;	Mismatches 11; Indels 0; Gaps 0;
QY	7	GDITQPTGKTVFNAPYDDKHTYHIKVINSSARRIGYIKTTNMKRLGVDP	PCGVLDPK	66	
ddb	1	GDITQPGSKTVFNAPYDDKHTYHIKITNAGRRRIGWAIKTTNMKRLGVDP	PCGVLDPK	60	
QY	67	AVLLAVSCDAFAFGQETNNDRITVENTPTDGAARQFRFEWQGG	113		
Db	61	NVLMAVSCDTEAAREINNDRIETWNTPDGAAKQFRFEWQGG	107		

RESULT
Q9GNV9

Query Match	72.48;	Score 494;	DB 5;	Length 107;
Best Local Similarity	82.28;	Pred. No. 1.1e-45;		

ID	Q9GNV9	PRELIMINARY;	PRT;	107 AA.
AC	Q9GNV9;			
DC	01-MAR-2001 (TrEMBLrel. 16, Created)			
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	MAJOR SPERM PROTEIN (FRAGMENT).			
GN	MSP.			
OS	Onchocerca volvulus.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;			
OC	Onchocercidae; Onchocera.			
OX	NCBI_TaxID=6282;			
RN	[1]			
RF	SEQUENCE FROM N.A.			
RC	STRAIN=OVNODBRA2C4 ;			
RX	MEDLINE=12321180; PubMed=11428336;			
RA	Morales Hojas R., Post R.J.;			
RT	"Regional genetic variation in the major sperm protein genes of			
RT	Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea)."			
RL	Int. J. Parasitol. 30:1459-1465(2000).			
DR	EMBL: AJ404207; CAC30741.1; -			
DR	HSSP: P27439; 1MSP.			
DR	InterPro: IPR000535; MSP_domain.			
DR	Pham; PF00635; MSP_domain; 1.			
FT	NON_TER	1		
FT	NON_TER	107		
FT	SEQUENCE	107 AA; 12017 MW;		OC5E932C86E5DF07 CRC64.

	Query Match	72.7%	Score 496;	DB 5;	Length 107;
	Best Local Similarity	82.2%	Pred. No. 6.Se-46;		
	Matches 88;	Conservative	8;	Mismatches 11;	Indels 0; Gaps
QY	7	GDITQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPDPCGVLDPKR	66		
Db	1	GDHTOPGSKIIVFNAPYDDKHTYHIKTIINAGRRIGAIKTTNMKRLGVDPDPCGVLDPKR	60		
QY	67	AVLLAVSCDAFAFGQEDTNNDRIIVENTNTPDGAARQFRFEWFGDG	113		
Db	61	NVLMAVSCDTFATDINNDRIIVENTNTPDGAARQFRFEWFGDG	107		

RESULT	14	
Q9GNW0		
ID	Q9GNW0	PRELIMINARY; PRT; 107 AA.
AC	Q9GNW0;	
DT	01-VAR-2001 (TremBirel. 16, Created)	
DT	01-VAR-2001 (TremBirel. 16, Last sequence update)	
DT	01-DEC-2001 (TremBirel. 19, Last annotation update)	
DE	MAJOR SPERM PROTEIN (FRAGMENT).	
GN	MSP.	
OS	Onchocerca volvulus.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;	
OC	Onchocercidae; Onchocerca.	
OX	NCBI_TaxID=6282;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=OVNODERA2C3 ;	
RA	MEDLINE=21321180; PubMed=11428336;	
RX	Morales Hojas R., Post R.J.;	
RT	"Regional genetic variation in the major sperm protein genes of	
RT	Onchocerca volvulus and Mansonella ozzardi (Nematoda : Filarioidea).";	
RL	Int. J. Parasitol. 30:1459-1465(2000).	
DR	EMBL; AJ404206; CAC20740.1; -.	
DR	HSSP; P27439; IMSP.	
DR	InterPro; IPR000535; MSP_domain.	
DR	Pfam; PF00635; MSP_domain; 1.	
FT	NON_TER 1	
FT	NON_TER 107	
SEQ	SEQUENCE 107 AA; 12018 MW;	OC5E932130E5DF07 CRC64:

Mon Sep 23 07:34:04 2002

Matches 88; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 7 GDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCGVLDPK 66
Db 1 GDIHTQPGSKIIVFNAPYDDKHTYHIKITNAGGRIGIWAIKTTNMKRLGVDPGCGVLDPK 60
QY 67 AVLLAVSCDAFAFGQEDTNDRIITVEWTNTPDGAARQFRREWFQGDG 113
Db 61 NVLMVSCDTFDTATREDINNDRIITVEWTNTPDGAARQFRREWFQGDG 107

RESULT 15
Q9GNI0 PRELIMINARY; PRT; 107 AA.
AC Q9GNI0;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE MAJOR SPERM PROTEIN (FRAGMENT).
GN MSP.
OS Mansonella ozzardi.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Mansonella.
OX NCBI_TaxID=122354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VARIOUS STRAINS;
RX MEDLINE=21321180; PubMed=11428336;
RA Morales Hojas R., Post R.J.;
RT "Regional genetic variation in the major sperm protein genes of
Onchocerca volvulus and Mansonella ozzardi (Nematoda: Filarioidea).";
RL Int. J. Parasitol. 30:1459-1465(2000).
DR EMBL; AJ404225; CAC20724.1; -
DR EMBL; AJ404210; CAC20709.1; -
DR EMBL; AJ404211; CAC20710.1; -
DR EMBL; AJ404212; CAC20711.1; -
DR EMBL; AJ404213; CAC20712.1; -
DR EMBL; AJ404215; CAC20714.1; -
DR EMBL; AJ404219; CAC20718.1; -
DR EMBL; AJ404221; CAC20720.1; -
DR EMBL; AJ404222; CAC20721.1; -
DR HSSP; P27439; IMSP.
DR InterPro; IPR000535; MSP_domain.
DR Pfam; PF00635; MSP_domain; 1.
FT NON_TER 1
FT NON_TER 107
SQ SEQUENCE 107 AA; 12017 MW; 1A0BD02C86E5C61E CRC64;

Query Match 72.3%; Score 493; DB 5; Length 107;
Best Local Similarity 81.3%; Pred. No. 1.4e-45;
Matches 87; Conservative 9; Mismatches 11; Indels 0; Gaps 0;

QY 7 GDIQTQPGTKIVFNAPYDDKHTYHIKVINSSARRIGYGIKTTNMKRLGVDPGCGVLDPK 66
Db 1 GDIHTQPGSKIIVFNAPYDDKHTYHIKITNAGGRIGIWAIKTTNMKRLGVDPGCGVLDPK 60
QY 67 AVLLAVSCDAFAFGQEDTNDRIITVEWTNTPDGAARQFRREWFQGDG 113
Db 61 NVLMVSCDTFDTATREDINNDRIITVEWTNTPDGAARQFRREWFQGDG 107

Search completed: September 20, 2002, 09:25:03
Job time: 141 sec

